

How to design a governable digital health ecosystem

Jessica Morley,¹ L Floridi^{1,2}

¹Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford, OX1 3JS, United Kingdom

²The Alan Turing Institute, 96 Euston Road, London, NW1 2DB, United Kingdom.

Correspondence email: jessica.morley@kellogg.ox.ac.uk

Statement of Contribution: JM is the main author of this article, to which LF has contributed.

Abstract

It has been suggested that to overcome the challenges facing the UK's National Health Service (NHS) of an ageing population and reduced available funding, the NHS should be transformed into a more informationally mature and heterogeneous organisation, reliant on data-based and algorithmically-driven interactions between human, artificial, and hybrid (semi-artificial) agents. This transformation process would offer significant benefit to patients, clinicians, and the overall system, but it would also rely on a fundamental transformation of the healthcare system in a way that poses significant governance challenges. In this article, we argue that a fruitful way to overcome these challenges is by adopting a pro-ethical approach to design that analyses the system as a whole, keeps society-in-the-loop throughout the process, and distributes responsibility evenly across all nodes in the system.

Keywords

Artificial Intelligence; Data-driven Services; Healthcare; National Health Service (NHS); Pro-ethical Design.

1. Introduction

In 2018, the UK's National Health Service (NHS) turned 70. Built on the guiding principles of belonging to all (Department of Health and Social Care, 2015), and providing free high-quality care at the point of need, the NHS is not only the UK's largest institution but also one of its most revered. However, it is also starting to show its age. In December 2018, the total number of patients waiting six weeks or longer from referral for one of the 15 key diagnostic tests was the highest in over ten years (NHS England, 2019a). And at the end of 2017, approximately 23% of all deaths in the UK were considered avoidable (Office for National Statistics, 2019). Given the demographic and funding changes challenging the NHS, this decline in performance is perhaps understandable. However, it is also morally unacceptable, when the overall result is loss in quality of life at best and loss of life at worst. What can be done?

Setting aside the need to tackle the wider socio-economic factors (de Freitas & Martin, 2015) that constrain people's ability to improve their health (Lewis, 2006),¹ the answer may well lie in making the NHS more informationally mature (Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2017), and able to capitalise on the opportunities presented by digital, data and artificial intelligence (particularly the use of machine learning (ML) (Ashrafian, Darzi, & Athanasiou, 2015)). Indeed, companies including DeepMind, Sensyne Health and Kheiron Medical Technologies have already started to make bold claims about the ability of their data-driven solutions to improve significantly outcomes for patients, and to reduce the burdens on the system. Systems that support clinical decision are in widespread use and have had a significant impact on safe prescribing, guidance adherence, prognostic screening, and risk scoring (Challen et al., 2019). Diagnostic screening has also seen considerable attention from the ML community. Today, AI systems can estimate bone age, diagnose retinal disease, or quantify cardiac risk with greater consistency, speed, and reproducibility than humans (He et al., 2019). These results are particularly exciting to healthcare policymakers when one considers that they are all instances of automating tasks that are

¹ It is vital that factors such as the cost of healthy food, varying levels of education regarding nutrition and exercise and issues with access to service are tackled if health outcomes are to be improved for the entire population, but for the purposes of this discussion a detailed investigation of the influence of these elements is not necessary.

not especially difficult (Floridi, 2019c). It is predicted that further benefits will be released in the future as the NHS becomes a learning health care system (Faden et al., 2013) (Celi et al., 2016).

In this envisaged data-based and algorithmically-driven system, there would be a continuous circulation of data (including genomic, phenotypic, behavioural, and environmental) from a range of sources (including wearables, images, social media, geo-spatial information (Nag et al., 2018) between patients, clinicians and care systems. In such a closed loop system, actionable advice, also based on AI, could be given to people before problems become significant (Nag, Pandey, Oh, & Jain, 2017), and demand for services could be predicted in advance, creating the opportunity, hypothetically, to reduce very significantly the percentage of avoidable deaths.

Such opportunities are not to be ignored, but capitalising on them is challenging. This is because the opportunities are not created by the technologies *per se* but by their ability to re-ontologise (fundamentally re-engineer the intrinsic nature of) the ways in which health care is delivered in the NHS by coupling, re-coupling and de-coupling different parts of the system (Floridi, 2017a). For example:

- Coupling: patients and their data are so strictly and interchangeably linked that the patients *are* their genetic profiles, latest blood results, personal information, allergies etc. (Floridi, 2017a). What the legislation calls ‘data subjects’ become “data patients”;
- Re-Coupling: research and practice have been sharply divided since the publication of the National Commission for the Protection of Human Subjects in the 1970s, but in the digital scenario described above, they are re-joined as one and the same again (Petrini, 2015) (Faden et al., 2013);
- De-Coupling: presence of Health Care Provider (HCP) and location of Patient become independent, for example because of the introduction of online consultations (NHS England, 2019b).

This process is changing the sociotechnical nature of the NHS. Interactions now occur between human, artificial and hybrid agents, and data is far more easily managed, duplicated, stored, mined and distributed (Turilli & Floridi, 2009). As a result, social, ethical and professional norms are being challenged by new issues related to fairness, accountability and transparency (Lepri, Oliver, Letouzé, Pentland, & Vinck, 2018). In short, increasing the NHS's reliance on 'information' increases its affordances (Floridi, 2017a) but also presents regulators and policymakers with considerable challenges related to its governance.²

Clearly, there is a need to take a proactive, 'digital ethics' approach to these challenges, so that the transition from an evaluation of what is morally good to what is politically feasible and legally enforceable (Floridi, 2018) happens before ethical mistakes lead to social rejection, and leave the NHS unable to benefit from the so-called 'dual advantage' (where opportunities are capitalised on and risks mitigated (Floridi et al., 2018)) of an ethical approach to governance (Floridi et al., 2018). The purpose of the following pages is to contribute to the development of a proactive approach. In section 2, we discuss the need to analyse the governance challenges posed by the digitisation of the NHS from the perspective of the whole system, if appropriate mitigating strategies are to be identified. In section 3, we provide a series of illustrative examples of pro-ethical governance decisions that could be used to meet four necessary conditions for successfully governed digitisation: data access, data protection, accountability, and evidence. In section 4, we reiterate that the examples given in section 3 are only illustrative of the art of the *possible*, not the art of the socially *preferable* and thus cannot be guaranteed to meet the fifth essential condition of successfully governed digital transformation, which is trust. For this, we argue that a mechanism for keeping society-in-the-loop during the elicitation of design requirements is essential. Finally, we conclude that the likelihood of such an approach succeeding remains uncertain, but this does

² Governance is taken to encompass regulation, ethics (of data, algorithms and practice (Floridi & Taddeo, 2016) and governance (in terms of policies, procedures and standards for the proper development, use and management of data-driven health and care tech (Floridi, 2018).

not mean that we should not try because, if we do not, we run the risk of the NHS ceasing to be for everyone, as it is designed to be.

2. A Systemic Approach

Reflection on the ethical implications of medical intervention has been a feature of delivering medical care since antiquity (Mann, Savulescu, & Sahakian, 2016). Medical practitioners' promise 'to do no harm' to their individual patients, and the bioethical principles of *beneficence*, *non-maleficence*, *justice*, and *autonomy* (Beauchamp & Childress, 2013) are well established in the medical literature. They have recently been adopted (along with the new principle of 'explicability') by the 'ethical Artificial Intelligence (AI)' community (Floridi et al., 2018a) in one of many attempts to encourage the development of algorithmic systems that are *fair*, *accountable*, and *transparent* (Lepri et al., 2018). This coming together of bioethics and AI ethics—coherent with a broader, patient-oriented approach to moral philosophy (Floridi, 2013b)—is essential, given the vast array of harms related to the potential for AI to replicate or exacerbate bias and behave in unexpected risky ways; to alter the interaction between patients and healthcare professionals; to change people's perception of their responsibility in managing their own body (Verbeek, 2009); and to use hugely personal information to manipulate patient behaviour without them realising it (Berdichevsky & Neuenschwander, 1999). However, this focus on the bioethical principles has prompted governance responses, in terms of policy and regulation, that focus solely on *individual* level impacts. For example, the NHS's Code of Conduct for Data-Driven Health and Care Technology (Department of Health and Social Care, 2019)—which rests on the Nuffield Council of Bioethics principles for data initiatives (respect for persons, respect for human rights, participation and accountability (Nuffield Council of Bioethics, 2015))—asks developers to consider their 'specific' user when carrying out tasks, such as a data protection impact assessment, data minimisation processes, and evaluation of evidence of effectiveness. However, it gives no guidance to

commissioners on how to assess the impact of introducing an algorithmic service at a group level (Taylor, Floridi, & van der Sloot, 2017).

Although vital, this exclusive focus on an individual level fails to recognise the risks that are associated with the fact that data related to healthcare of whole groups of people and not just individuals are now circulating outside of NHS boundaries, shared with third parties for research and commercial purposes (Garattini, Raffle, Aisyah, Sartain, & Kozlakidis, 2019), connecting personal, provider and population health information in complex feedback loops that exist at many levels (Flahault et al., 2017). In order to manage effectively these risks, and ensure that the NHS as a whole can both benefit from the dual advantage of ethical governance (in terms of risk management and opportunity strategy) and keep to its commitment of belonging to all, it is necessary to take into account a broader set of observable variables than recognised or protected in the above outlined policies, by adopting a different Level of Abstraction (LoA)³ (Floridi, 2008). The appropriate LoA is one that (a) looks at the systems level and considers the entire human, social and organisational infrastructure in which data-driven health and care technologies are being embedded (Macrae, 2019) and (b) involves public voices (Gonzalez-Polledo, 2018), so that the societal implications become clear (O'Doherty et al., 2016). A systems-level analysis, as set out below, will highlight the emergent impacts on fairness, accountability and transparency (Lepri et al., 2018) that result from the interaction between connected system components (Rebhan, 2017), and produce a more holistic understanding of the governance challenges facing an informationally-maturing NHS (Crawford & Calo, 2016) than possible when analysed at the individual patient LoA.

³ A level of abstraction can be imagined as an interface that enables one to observe some aspects of a system analysed, while making other aspects opaque or indeed invisible. For example, one may analyse a house at the LoA of a buyer, of an architect, of a city planner, of a plumber, and so on. LoAs are common in computer science, where systems are described at different LoAs (computational, hardware, user-centred etc.). LoAs can be combined in more complex sets, and can be, but are not necessarily always, hierarchical.

2.1 Fairness at the systems level

An informationally-mature NHS will provide better care for individuals about whom there is more information from which it can learn. Typically, those who generate less information (data) about themselves are conceptualised as those with less access, for example those who do not have a smartphone. However, a recent mixed-methods study by Powell & Deetjen (2019) found that there are at least six types of health information seekers (and therefore data generators): learners, pragmatists, sceptics, worriers, delegators, and adigitals. Each of these different groups will input different amounts of personal information into the system's 'training data', which will be used to create predictive models (Chiu & Hripcsak, 2017) (profiles) that will form the basis of decisions about, for example, allocation of resource at the systems level. Unless carefully overseen—through the use of techniques such as 'data nutrition labelling' (Holland, Hosny, Newman, Joseph, & Chmielinski, 2018), model cards (Mitchell et al., 2019) or datasheets for datasets (Gebru et al., 2018)—this process will make it impossible for the agents (human or otherwise) running the NHS to guarantee procedural regularity (Kroll et al., 2017). And without this guarantee, this process could lead to discriminatory practices and outcomes (Lepri et al., 2018). In the worst-case-scenario, this could render entire groups of the population 'invisible' to the system.

Similar disproportionate effects can result from the way the 'system' incentivises the spread of innovation, including data-driven innovation. NHS Trusts that are performing well—in terms of key performance indicators such as A&E wait time—receive more funding by being rewarded for meeting targets and by avoiding fines. These Trusts, such as Moorfields Eye Hospital, are then more able to divert funding away from frontline care costs to invest in the successful implementation and ongoing process improvement of as-of-yet 'unproven' technologies such as AI (He et al., 2019). If the implementation proves successful and improves patient outcomes, the Trust will likely receive further financial incentivisation through schemes such as the 'Global Digital Exemplars' programme (NHS England, 2019b) and, potentially, profit from intellectual property rights. This is leading to a clear polarisation, with already high-performing Trusts

increasingly benefitting from the central drive to ‘digitise’ the NHS and forces related to cumulative advantage, whilst underperforming Trusts, whose patients would likely see *more* benefit from the implementation of data-driven technologies, increasingly missing the opportunity to reap such benefits and so becoming cumulatively more disadvantaged (DiMaggio & Garip, 2012).

2.2 Accountability and transparency at the systems level

An accountable decision-maker is one that is able to provide its decision-subjects with reasons and explanations for the design and operationalisation of its decision-making system, so that the decision-subject can judge whether this justification is adequate, and if not, sanction the decision-maker (Binns, 2018). In this context, the NHS is the decision-maker and its users are the decision-subjects. At the systems LoA, this means that users should be able to understand and question why decisions were made about, for example, funding allocations, particularly as this funding is public money, such as why treatment for one condition is available on the NHS and not another, and how and/or why their data were involved in making these decisions. However, it is increasingly difficult for the system to be able to provide similar explanations. This is because it has become considerably more challenging for the system to understand the reliability of the data or where the data were generated, as data have come to flow more easily from data collectors, to data aggregators, to analysers, (Garattini et al., 2019) not all of whom belong to the NHS, and some of whom may be under no obligation to tell policymakers or implementers how the data were collected, aggregated, stored, or processed.

This lack of oversight and transparency is a direct result of the NHS’s ‘information revolution’, which has forced it to become a more heterogeneous organisation (Turilli & Floridi, 2009) or set of organisations. For instance, Babylon, a private third party, provides the Skype Consultation service ‘GP at Hand’ for some NHS GP practices in London (GP at hand, 2019). This organisational complexity has effectively rendered much of the NHS back-end decision-making systems ‘black boxes.’ This is concerning (Watson et al., 2019) because, if something goes

wrong with the allocation of resource (for example, if there is insufficient staff placed in an emergency unit, and this results in fatalities), or with informational security (such as WannaCry (Department of Health and Social Care, 2018)) or patient data privacy, it is not clear enough which ‘nodes’ in the system (Floridi, 2016a) made the decision in order to hold it accountable. Indeed, it may take much longer for problems felt at the point of service delivery to become visible at the point of service provision. This makes it considerably harder for users to trust the system (Sterckx, Rakic, Cockbain, & Borry, 2016), as it becomes increasingly reliant on processes that they cannot understand.

3. A proactive approach to ethical governance

When the ‘problem’ of the digital transformation of the NHS is so framed, it becomes easier to understand how every element of the digital ecosystem, even the seemingly innocuous—such as naming the NHS’s digital transformation portfolio ‘empower the person (NHS England, 2018)’ (Morley and Floridi, 2019)—can be seen as a moral enabler or a moral hindrance (Floridi, 2013b)). It also becomes clearer how each element of the system can be ‘pro-ethically’⁴ designed to protect the values, principles and ethics that we think are fundamental to the NHS (Floridi, 2017b). This requires good ethical governance (both soft and hard) that translates between principles and practice (Winfield & Jirotko, 2018) in a way that addresses the challenges outlined and enables digital health innovation to produce better outcomes for all (Vayena, Tobias, Afua, & Allesandro, 2018).

⁴ The difference between ethics by design and pro-ethical design is the following: *ethics by design* can be paternalistic in ways that constrain the choices of agents, because it makes some options less easily available or not at all; instead, *pro-ethical design* still forces agents to make choices, but this time the nudge is less paternalistic because it does not preclude a course of action but requires agents to make up their mind about it. A simple example can clarify the difference. A speed camera is a form of nudging (drivers should respect the speed limits) but it is pro-ethical insofar as it leaves to the drivers the freedom to choose to pay a ticket, for example in case of an emergency. On the contrary, in terms of ethics by design, speed bumps are a different kind of traffic calming measure designed to slow down vehicles and improve safety. They may seem like a good idea, but they involve a physical alteration of the road, which is permanent and leaves no real choice to the driver. This means that emergency vehicles, such as a medical ambulance, a police car, or a fire engine, must also slow down, even when responding to an emergency.

According to Vayena et al. (2018) such governance should meet five conditions in order to ensure all the ethical and policy challenges posed by data-driven healthcare are covered: (1) data access; (2) data protection; (3) accountability; (4) evidence; and (5) trust. We shall address the condition of trust in section 4. In the rest of this section, we shall illustrate how governance decisions can be made to ensure that the first four of these conditions are met by collectively tackling the issues of confidentiality and consent for the public good (see 3.1); providing data protection legislation that enables individuals to ensure their data is *working for* them even without their awareness of it (see 3.2); creating a regulation as a service pathway (see 3.3); and investing in safe environments for experimentation (see 3.4). These clarifications are designed to show the art of the possible, but it should be remembered that, for each condition, the ‘how-to meet x’ question is an open one that may be ‘solved’ in more than one way (Floridi, 2017b).

3.1 Data access: collectively tackle the issues of confidentiality and consent for the public good

To unlock the benefits of data-driven healthcare, large datasets are needed for initial training, ongoing training, validation, and refinement of data-driven solutions (He et al., 2019). Many data sources exist, including population biobanks, cohort studies, genome databases, clinical and public health records, direct-to-consumer genetic test data, social media, and data collected from fitness trackers, health apps and biometric sensors (O’Doherty et al., 2016). Yet it can be phenomenally difficult for researchers to access these data, resulting in significant opportunity costs for the system (Watson et al., 2019). Partly this is to do with a lack of data standardisation and resulting issues with interoperability (He et al., 2019). However, this is improving with an increased recognition across the system that data need to be FAIR (findable, accessible, interoperable and reusable) (Kemper & Kolkman, 2018) and the mandating of specific data standards for all digital services in use by the NHS by NHSX from July 2019 (NHSx, 2019). A far more complicated challenge is answering when, by whom, why, where (for example by a commercial or non-

commercial company or for profit or not-for-profit purposes) and how health data can be accessed (Floridi et al., 2018).

Typically, these questions have been answered by addressing first another question concerning what uses the relevant patient has provided informed consent for. However, it is now well recognised that attaining traditional informed consent is no longer feasible in many data-driven contexts where the specific objectives and risks of research are not known *a priori* (Woolley, 2019). Given that the requirement for informed consent can act as a significant barrier to the generation of biomedical knowledge that can save lives (Mann et al., 2016), European Data Protection Law does allow for the processing of personal health data without consent when it is in the ‘public interest’ to do so (Quinn, 2017). However, concerns about breaching the common law of confidentiality, effectively enshrined in the Hippocratic Oath, have left many researchers reluctant to take any risk (Mann et al., 2016).

In response, several authors have argued that we should move to a model of broad (or meta) consent, based on the ‘duty of easy rescue’ where it is presumed that—when the benefits of sharing health data in terms of the generation of scientific knowledge, significantly outweigh the potential harms to individuals—individuals have consented to its use as a public good (Mann et al., 2016, Ballantyne & Schaefer, 2018, Floridi et al., 2018b). Indeed, in 2014 the Global Alliance for Genomics and Health (GA4GH) created the framework for Responsible Sharing of Genomics and Health-Related Data with the explicit goal of moving towards a vision of data-intensive research for the common good as an expression of the fundamental human right to benefit from science (Knoppers & Thorogood, 2017). Yet this argument has found it hard to gain traction more broadly, at least partly because it has been presumed that, in order for such an approach to be ethically viable, the definition of a public good or science must be agreed by all. Yet this is unlikely to ever succeed as people disagree about values and what defines knowledge (Ploug & Holm, 2016), making it impossible to come up with a set of universally acceptable moral rules (Binns, 2018) of when it is acceptable to presume broad consent. A different approach may be more

successful and arguably the pro-ethical design approach we are advocating (Floridi, 2016b) could contribute to ensure that the public is engaged appropriately and agrees collaboratively (Georgiou et al., 2018) on the current ‘red-lines’ (or scenarios) when data-sharing must *never* be based on presumed or broad consent (more on this later). Crucially, the guiding principle of this process must be toleration (Floridi, 2016b) (where an act of toleration is defined as an agent’s intentional and principled refraining from interfering with an opposed other (or their behaviour) in situations of diversity where the agent believes she has the power to interfere (Cohen, 2004)), so that the system can cope with people who decide to opt out of being part of the data sharing ecosystem entirely.

3.2 Data protection: enable competition to ensure fair return on data investment

Since its introduction in May 2018, the European General Data Protection Regulation has been, to a greater or lesser extent, heralded as a solution for ensuring the protection of an individual’s data. It is, however, by no means perfect (Townend, 2018) and, given its focus on anonymisation, in the case of personalised medicine, it may be already out-of-date.

If the benefits outlined in the introduction of a truly personalised data-driven health and care system are to be realised, it will be necessary to develop a regulatory framework for data protection that enables the integration of data related to multiple ‘omics’ (Okun & Wicks, 2018) in a way that is respectful of the subjective nature of privacy (Alshammari & Simpson, 2017). This might seem an impossible task, and yet there is one potentially simple policy change that could support this and has been provided by the financial services sector: enable regulated aggregation so that individuals, groups and populations get a fair return on the investment of their data. This may work in the following way.

Payment Services Directive 2 (PSD2) (Payment services (PSD 2) - Directive (EU) 2015/2366), which became enforceable in January 2018, requires all retail financial services providers, upon the request of a customer, to open up their payments infrastructure and customer

data assets to a third party so that said third party can provide information services (most commonly *data aggregation* services) for the individual customer in question. It was introduced partly in response to the lack of competition in the retail banking sector when other previous initiatives, such as the current account switching service, failed to incentivise incumbent retail banks to compete for business in a way that led to better financial outcomes for individual customers.

By making it a legal requirement for bank *A* to open up its data at the request of bank *B* on behalf of customer *C*, the legislation enables *B* to compete for the custom of *C* by providing them with a better service than was provided by *A*. This not only ensures *C* gets the best service for them, but it also drives up the quality of services for all who currently use retail banking services (note, however that it does little to improve the outcomes for the unbanked community) by forcing the incumbents to compete for business rather than complacently relying on the inertia of their existing customers.

The same logic can work in the healthcare sector by enabling individuals, be they patients or citizens, to ensure that their data is *working for* them, rather than merely ‘empowering’ them. The difference might not be immediately apparent, but it is crucial. In the current climate of *empowerment*, as we have argued previously (Morley and Floridi, 2019) patient *P* is presented with their health data (e.g. GP record, or statistics from Garmin Connect) by a digital health provider *A* with the expectation that *P* will translate reflection on the data presented by *A* into actions that move them towards pre-conceived ‘targets’ (e.g. number of steps, calorie deficit or optimum resting heart rate) – even if those targets may be impossible or inappropriate for them to meet. In this scenario, there is no incentive for *A* to provide services that *genuinely improve* the health outcomes of patient *P*, and the only return on the data investment made by *P* is an increased responsibility for managing their own health, irrespective of whether or not this is feasible. This means that, even if that individual’s data are protected in a quantitative sense (i.e. there are appropriate security measures in place), they are not protected in a qualitative sense because they have delivered no benefit to *P*. Instead, the data have potentially been used solely for commercial

gain by A . In the comparative 'PSD2-enabled' scenario, if P does not feel as though they are getting a fair return on their current data investment from A —for example they are struggling to get a diagnosis, cannot access the services they need, or they feel as though their data is being exploited for inappropriate purposes— P can choose to share their data with B . B represents an aggregator that has stricter policies about use of data for pure commercial gain and is better able to act on the behalf of the individual by, for example, recommending health interventions tailored to P 's circumstances, making suggested diagnosis or referral, or requesting an appropriate prescription from P 's clinician. In short, the legislation would enable digital health tools (DHTs) to act as *digital companions* and encourage would-be data aggregators to compete based on the provision of services that enable individuals, groups and populations to meet appropriately tailored health outcomes.

Importantly, as with the case in the financial services sector, this policy intervention not only could improve the outcomes for data-rich and digitally-engaged individuals, it could also be used to enable the development of a learning healthcare system (Nag et al., 2017) that could benefit *all* users. This is because it would enable those using DHTs for personal benefit also to request that their data generated by third parties, are provided back to the NHS. In this way, the NHS could gain greater oversight of the wider digital health data economy and obtain richer data, both in terms of quality and quantity, than it is currently able to generate by itself. This would at least partially overcome some of the issues highlighted in section 2.2 and, if the right supporting policies were in place, enable the NHS to make better informed decisions about the provision of more tailored services (both digital and physical) based on the patient profiles generated by the aggregation of the data provided through this mechanism, rather than those provided by existing clinical trials data or GP records, which are inherently limited and often biased.

The previous analysis is not meant to suggest that achieving these outcomes will be easy. The proposal is complex and presents a significant multi-layered pro-ethical design challenge. For example, to ensure that everyone is able to benefit equally from the promise we have set out of

user-controlled data sharing and aggregation, there is a need to invest in the following (as a minimum):

1. Mandatory data standards and open APIs to ensure interoperability
2. New technical solutions for eliciting data-sharing preferences by presenting terms and conditions in more accessible formats, such as video or RIB, and accepting consent in forms other than written signature, for example gesture (Butterworth, 2018)
3. Privacy preserving techniques, such as homomorphic encryption, differential privacy, and federated learning
4. Training of frontline staff (for example General Practitioners and practice managers) to enable them to provide an offline user-journey that gives those who do not want to engage with the healthcare system digitally equal control over what happens to their data
5. The NHS's data science capability so it is able to benefit from the new sources of data in the manner described
6. An accreditation process for DHT providers wishing to become health data aggregators to process medical data (e.g. official medical records) as well as 'wellness' data (e.g. FitBit data). The minimum viable accreditation process should cover all the areas of the NHS's Digital Assessment Questionnaire: clinical safety, data protection, security, usability and accessibility, interoperability and technical stability and most crucially intended outcomes. A return on data investment can only be considered 'fair', and therefore acceptable in the NHS, if the intended outcome of the aggregation is a genuine improvement of the user's (or a group of users') health outcomes. A would-be aggregator that simply wants access for the purpose of providing them with targeted drugs advertising, for example, should not be considered acceptable. The onus of deciding which aggregators are 'trustworthy' should not sit with individual citizens.

If these challenges can be effectively tackled, such a combination of policy, technical and design requirements could result in significant improvements to the health of individuals and to the quality of healthcare as a whole. If not, the creation of the policy framework will most likely hinder rather than help the creation of a pro-ethically designed digital healthcare system, leading to a scenario where the potential unequal outcomes described in section 2.1 are exacerbated.

3.3 Accountability: create a 'regulation as a service' pathway

Currently, the governance framework (both in terms of hard and soft governance) for the development, deployment and use of data-driven health technology is convoluted, making it hard to navigate, discouraging regulatory compliance, and potentially undermining public trust. This is primarily the result of a culture of caution (Sethi & Laurie, 2013). What is needed—instead of the collection of regulations, policies and standards that have been developed in a relatively *ad hoc* fashion—is a coordinated effort from the centre of the healthcare system to move regulation away from being something that technology providers interact with at specific points-in-time towards 'regulation as an enabling service'. Such a service should cover the entire development pathway of data-driven technologies, as illustrated in figure 1 adapted from the ICO's AI Auditing Framework's depiction of the AI lifecycle (Binns, 2018), and should include mechanisms for protecting patient safety (He et al., 2019) and ensuring transparency, accuracy and interpretability (Guidotti et al., 2018) both *ex ante* and *ex post* (Reed, 2018), as data-driven products continuously cycle from:

- (1) development: when regulations regarding research ethics and data access are needed and compliance is supported by newer techniques, such as federated learning (Bonawitz et al., 2019)); to
- (2) deployment: when regulations related to validation of results and data protection are needed; and to

- (3) use: when regulations related to continuous audit (Binns, 2018), and reporting of safety issues (Fong et al., 2018) are required.

Importantly, the regulatory building blocks that make up this pathway must meet two conditions. First, they must be proportionate to risk, so that the regulatory burden does not become disproportionate to the risks posed to a patient (Sethi & Laurie, 2013). Second, they must be put together in such a way that responsibility is identifiably distributed across different levels of the system, so that it is possible to trace back the cause of the result in order for it to be corrected or recreated (depending on whether the result was bad or good) (Floridi, 2016a). Technology corporations that produce data-driven health technology and algorithms must be held accountable (Owens & Cribb, 2017) for the types of tools they develop and how they ensure the design is trustworthy (Pak, Fink, Price, Bass, & Sturre, 2012), whilst the State must take responsibility for bringing together regulators, commissioners, technology providers and policymakers to achieve a common goal at the systems LoA.

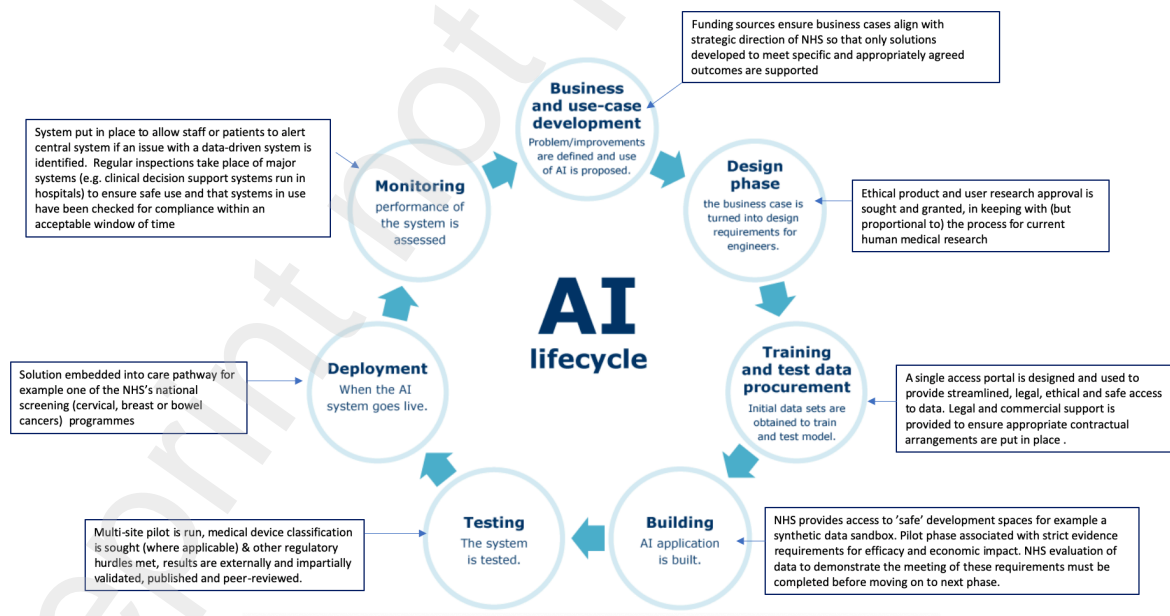


Figure 1: Illustrative example of 'Regulation as a Service' for data-driven health and care technologies

3.4 Evidence: invest in 'safe' environments for experimentation

In evidence-based medicine, the belief to be justified (by the evidence) concerns the effectiveness of a particular medical intervention for a specific patient (Eustace, 2018). This is no less true of data-driven medical interventions at the individual or system level than it is of 'traditional' medical interventions. To be certain of the effectiveness of an intervention, progress needs to be monitored over time to understand how interventions work at the systems level to cause health changes at the individual level (Michie, Fixsen, Grimshaw, & Eccles, 2009). This means that it is risky to implement the use of data-driven technologies into the complex health system, without any *a priori* knowledge of how they will develop over time, unless the impacts can be modelled and simulated ahead of time (Silverman, Hanrahan, Bharathy, Gordon, & Johnson, 2015).

The challenge is that, although the computing capacity to enable such modelling exists, the data that are required are difficult and expensive to generate and access (Krutzinna, Taddeo, & Floridi, 2018). Yet, this need not be the case. Currently, GP records of the deceased are passed to Primary Care Support England, where they are stored for ten years before being destroyed (NHS UK, 2018). Enabling the use of these data by conducting extensive stakeholder analysis, writing into policy the proposed ethical code for posthumous medical data donation (Krutzinna et al., 2018), and giving people the option, through the 'national data opt-out' programme, to opt out of 'donating' their data should they wish to do so, could be transformative. Appropriate tests would need to be used to ensure the data's completeness, quality and robustness but, providing the data contained within the records are sound, they could be anonymised and synthesised, to provide multiple layers of protection, and integrated with other government held records (e.g. care records, pensions information) to create a 'digital twin' of the healthcare system. This 'twin' could then be used as a sandbox environment for experimentation, creating the opportunity for data-driven technology to demonstrate effectiveness without any risk of patient harm, and enabling policymakers and commissioners to model the system-wide impact of the introduction of a specific

algorithmic or other data-driven system ahead of time. This would significantly improve the likelihood of successful implementation of data-driven health technologies.

4. **Keeping Society-in-the-Loop**

As stated previously, these outlined governance decisions should be seen as illustrative examples of *possible* means of meeting the conditions of data access, data protection, accountability and evidence, but not as the *only* means. This is because such governance decisions are definitely *possible*, and may even be *acceptable*, but there is no way of knowing a priori whether they are socially *preferable* (Floridi & Taddeo, 2016) and thus, they do not guarantee that the fifth essential condition of ‘trust’ will be met (Vayena et al., 2018). To ensure that this condition is met, the people who rely on the NHS to care for them and their loved ones, or to provide them with meaningful employment, must not be seen only as ‘stakeholders,’ (as they are in the above example of meta-consent), but as real interlocutors (able to participate in the shaping of the system), (Durante, 2014), ‘prosumers’ of a makers’ knowledge instead of a users’ knowledge (Floridi, 2019b). The process of making all pro-ethically designed governance decisions must not begin with the paternalistic and technocratic assumptions of the State, but with the elicitation of the requirements of these individuals. Doing this well is vital, because whether or not these design requirements are understood and correctly implemented, can make the difference between the failure of Care.Data (NHS England, 2013) and the relative success of the National Data Opt-Out (NHS Digital, 2019).

Gaining an understanding of these requirements necessitates asking different questions to account for conflicting viewpoints, complex cost/benefit comparisons, and the impact of cultural norms, institutional structures, and economic factors on people’s views on health system digitalisation (Balaram, Greenham, & Leonard, 2018). In short, to ensure that the designers of the digital NHS are able to work from an agreed set of requirements, it will be essential to develop what Rahwan (2018) refers to as a model for society-in-the-loop (SITL), by making use of research methods that keep people regularly engaged throughout the design process.

The research group *Understanding Patient Data* highlights several good practice examples of research conducted in this manner to elicit public attitudes to patient data (*Understanding Patient Data*, 2019). For example, the Academy of Medical Sciences commissioned Ipsos Mori (2018) to hold dialogue workshops with members of the public, patients and health care professionals on awareness, aspirations, expectations and concerns around uses of patient data in future technologies (Castell, Robinson, & Ashford, 2018). And Leeds Informatics Board commissioned Brainbox to hold conversations and facilitated workshops, analyse social media data and conduct journey mapping in a variety of different locations to explore attitudes to health data sharing (Fylon, 2015).

Crucially this participatory elicitation of SITL-requirements must be seen as a process, rather than a one-off act. This is because social values and norms shift over time and in different contexts, and therefore, what are seen as socially acceptable (and even preferable) uses and designs of technology and its governance will also shift. It will not be sufficient simply to ask society to outline ‘red-lines’ (as suggested earlier) for the use of digital technologies and data in the NHS once and assume that, provided design decisions do not breach these, the SITL-requirements have been met in perpetuity. Nor should it be assumed that eliciting these requirements will be straightforward. Principled disagreements are inevitable and should be welcomed as society grapples with the serious ethical considerations of the NHS increasingly relying on algorithmic decision-making.

Unfortunately, from the perspective of companies and governments this means that the costs of SITL-requirement elicitation cannot simply be borne as one-off revenue spend but must, instead, be factored into business-as-usual operating, or capital, costs. This is likely to be hard to justify when the benefits of ‘increased trust’ cannot be automatically translated into quantifiable return on financial investment. The responsibility of ensuring that this process is carried through cannot, therefore, sit with just one group of NHS designers (programmers, policymakers, regulators, healthcare professionals, academics etc.). Instead, building on the point made earlier about distributed regulatory responsibility, technology corporations that produce digital health

devices and algorithms must be held responsible (Owens & Cribb, 2017) for ensuring that the tools they develop meet the principle of autonomy, as well as the other relevant bioethical principles of beneficence and non-maleficence (Floridi *et al* 2018). The State must also take responsibility for creating ethically-aligned policies and for encouraging the system components to coordinate. Then researchers and charities must take responsibility for ensuring the knowledge of the users is heard. Finally, commissioners and healthcare practitioners must take responsibility for checking to see whether the technology services they are planning to procure, prescribe or use meet the socially-defined and ethically-aligned requirements (as well as any policy standards or regulatory requirements). Only in this scenario of distributed but clearly identifiable responsibility will it be possible to ensure the entire causal chain (which should really be seen as a continuous cycle of development → deployment → use) is aligned with societal *desiderata* (Lipton 2017).

5. Conclusion

With the NHS Long Term Plan underpinning the importance of technology in the future NHS to provide a step change in the way that it cares for citizens, it is clear that, whether digitisation will have an impact on the NHS, is no longer a relevant question. Instead, the relevant questions are those related to governance mechanisms and how these can be developed to ensure that the digital NHS is designed to distribute evenly the positive impacts and limit the negative impacts. Developing such mechanisms will be challenging but, as illustrated, not outside the realms of possibility or, indeed, preferability, if a pro-ethical approach to system design is taken.

Taking such an approach does require a willingness to live with uncertainty as the likelihood of any of the potentially implementable mechanisms succeeding in enabling the digital NHS system to capitalise on the dual advantage of ‘ethics’ remains an unknown (Vayena, et al., 2018). However, if this uncertainty can be embraced and discussed openly, it should not prevent us from trying. We should be mindful of the challenge but eager to understand and act on what

can be done safely and in a socially acceptable way so that the opportunities presented by data-driven technologies for the care of those who rely on the NHS are not wasted (Floridi, 2019a).

This understanding must be developed by keeping citizens and patients actively involved in every step of the governance design so that society is kept firmly in-the-loop (Rahwan, 2018). If the State fails to do this, it will produce a digital health ecosystem that supports only a very small number (Floridi, 2014) of algorithmically designed digital selves, whilst ignoring the health of those that fall outside of these categories. If this were to transpire, the NHS would cease to be for everyone, as it is designed to be, and no amount of governance will suffice to undo the damage. The risks and the opportunity costs associated with this outcome are too great to contemplate and so it is paramount that every node in the system is encouraged to amplify society's voices to ensure that the people for whom healthcare matters the most are seen as part of the solution to building trust in an informationally-mature NHS, rather than a hurdle to this trust that must be overcome (Aitken et al., 2019).

Funding

This work was partially supported by Privacy and Trust Stream - Social lead of the PETRAS Internet of Things research hub - PETRAS is funded by the UK Engineering and Physical Sciences Research Council (EPSRC), grant agreement no. EP/N023013/1. It was also partially supported by a Microsoft grant.

Bibliography

- Aitken, M., Tully, M. P., Porteous, C., Denegri, S., Cunningham-Burley, S., Banner, N., ... Willison, D. J. (2019). Consensus Statement on Public Involvement and Engagement with Data-Intensive Health Research. *International Journal of Population Data Science*, 4(1). <https://doi.org/10.23889/ijpds.v4i1.586>
- Alshammari, M., & Simpson, A. (2017). Towards a Principled Approach for Engineering Privacy by Design. In E. Schweighofer, H. Leitold, A. Mittrakas, & K. Rannenberg (Eds.), *Privacy Technologies and Policy* (Vol. 10518, pp. 161–177). https://doi.org/10.1007/978-3-319-67280-9_9
- Ashrafian, H., Darzi, A., & Athanasiou, T. (2015). A novel modification of the Turing test for artificial intelligence and robotics in healthcare: Modified Turing test for robotic healthcare. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 11(1), 38–43. <https://doi.org/10.1002/rcs.1570>
- Balaram, B., Greenham, T., & Leonard, J. (n.d.). *Artificial Intelligence: Real Public Engagement*. Retrieved from RSA website: https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf
- Ballantyne, A., & Schaefer, G. O. (2018). Consent and the ethical duty to participate in health data research. *Journal of Medical Ethics*, 44(6), 392–396. <https://doi.org/10.1136/medethics-2017-104550>
- Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics* (7th ed). New York: Oxford University Press.
- Berdichevsky, D., & Neuenschwander, E. (1999). Toward an ethics of persuasive technology. *Communications of the ACM*, 42(5), 51–58. <https://doi.org/10.1145/301353.301410>
- Binns, R. (2018). Algorithmic Accountability and Public Reason. *Philosophy & Technology*, 31(4), 543–556. <https://doi.org/10.1007/s13347-017-0263-5>
- Binns, R. (n.d.). An Overview of the Auditing Framework for Artificial Intelligence and its core components. Retrieved from ICO website: https://ai-auditingframework.blogspot.com/2019/03/an-overview-of-auditing-framework-for_26.html
- Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., ... Roselander, J. (2019). Towards Federated Learning at Scale: System Design. *ArXiv:1902.01046 [Cs, Stat]*. Retrieved from <http://arxiv.org/abs/1902.01046>
- Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. *Computer Law & Security Review*, 34(2), 257–268. <https://doi.org/10.1016/j.clsr.2018.01.004>
- Castell, S., Robinson, L., & Ashford, H. (2018). *Future data-driven technologies and the implications for use of patient data* (p. 44). Retrieved from Ipsos Mori website: <https://acmedsci.ac.uk/file-download/6616969>
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2017). Artificial Intelligence and the ‘Good Society’: The US, EU, and UK approach. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-017-9901-7>
- Celi, L. A., Davidzon, G., Johnson, A. E. W., Komorowski, M., Marshall, D. C., Nair, S. S., ... Stone, D. J. (2016). Bridging the health data divide. *Journal of Medical Internet Research*, 18(12). <https://doi.org/10.2196/jmir.6400>
- Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, 28(3), 231–237. <https://doi.org/10.1136/bmjqs-2018-008370>
- Chiu, P.-H., & Hripcsak, G. (2017). EHR-based phenotyping: Bulk learning and evaluation. *Journal of Biomedical Informatics*, 70, 35–51. <https://doi.org/10.1016/j.jbi.2017.04.009>
- Cohen, A. J. (2004). What Toleration Is. *Ethics*, 115(1), 68–95. <https://doi.org/10.1086/421982>
- Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, 538(7625), 311–313. <https://doi.org/10.1038/538311a>
- de Freitas, C., & Martin, G. (2015). Inclusive public participation in health: Policy, practice and theoretical contributions to promote the involvement of marginalised groups in healthcare. *Social Science & Medicine*, 135, 31–39. <https://doi.org/10.1016/j.socscimed.2015.04.019>
- Department of Health and Social Care. (2015). *The NHS Constitution for England*. Retrieved from <https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england>
- Department of Health and Social Care. (2018). *Lessons learned review of the WannaCry Ransomware Cyber Attack*. Retrieved from <https://www.england.nhs.uk/wp-content/uploads/2018/02/lessons-learned-review-wannacry-ransomware-cyber-attack-cio-review.pdf>
- Department of Health and Social Care. (2019, February 19). Code of conduct for data-driven health and care technology. Retrieved 15 April 2019, from GOV.UK website: <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology>
- DiMaggio, P., & Garip, F. (2012). Network Effects and Social Inequality. *Annual Review of Sociology*, 38(1), 93–118. <https://doi.org/10.1146/annurev.soc.012809.102545>
- Durante, M. (2014). *The Democratic Governance of Information Societies. A Critique to the Theory of Stakeholders* (Vol. 28).
- Effy, V., Tobias, H., Afua, A., & Alessandro, B. (2018). Digital health: Meeting the ethical and policy challenges. *Swiss Medical Weekly*, 148(3–4). <https://doi.org/10.4414/smww.2018.14571>
- European Commission. *Payment services (PSD 2) - Directive (EU) 2015/2366*.
- Eustace, S. (2018). Technology-induced bias in the theory of evidence-based medicine. *Journal of Evaluation in Clinical Practice*, 24(5), 945–949. <https://doi.org/10.1111/jep.12972>

- Faden, R. R., Kass, N. E., Goodman, S. N., Pronovost, P., Tunis, S., & Beauchamp, T. L. (2013). An Ethics Framework for a Learning Health Care System: *A Departure from Traditional Research Ethics and Clinical Ethics*. *Hastings Center Report*, 43(s1), S16–S27. <https://doi.org/10.1002/hast.134>
- Flahault, A., Geissbuhler, A., Guessous, I., Guérin, P., Bolon, I., Salathé, M., & Escher, G. (2017). Precision global health in the digital age. *Swiss Medical Weekly*, 147(1314). <https://doi.org/10.4414/smww.2017.14423>
- Floridi, L., Luetge, C., Pagallo, U., Schafer, B., Valcke, P., Vayena, E., ... Kalra, D. (2018). Key Ethical Challenges in the European Medical Information Framework. *Minds and Machines*, 1–17. <https://doi.org/10.1007/s11023-018-9467-4>
- Floridi, L. (2008). The Method of Levels of Abstraction. *Minds and Machines*, 18(3), 303–329. <https://doi.org/10.1007/s11023-008-9113-7>
- Floridi, L. (2013a). Distributed Morality in an Information Society. *Science and Engineering Ethics*, 19(3), 727–743. <https://doi.org/10.1007/s11948-012-9413-4>
- Floridi, L. (2013b). *The ethics of information*. Oxford: Oxford University Press.
- Floridi, L. (2014). *The 4th revolution: How the infosphere is reshaping human reality*. Oxford: Oxford Univ. Press.
- Floridi, L. (2016a). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), 20160112. <https://doi.org/10.1098/rsta.2016.0112>
- Floridi, L. (2016b). Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration. *Science and Engineering Ethics*, 22(6), 1669–1688. <https://doi.org/10.1007/s11948-015-9733-2>
- Floridi, L. (2017a). Digital's Cleaving Power and Its Consequences. *Philosophy & Technology*, 30(2), 123–129. <https://doi.org/10.1007/s13347-017-0259-1>
- Floridi, L. (2017b). The Logic of Design as a Conceptual Logic of Information. *Minds and Machines*, 27(3), 495–519. <https://doi.org/10.1007/s11023-017-9438-1>
- Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 376(2133). <https://doi.org/10.1098/rsta.2018.0081>
- Floridi, L. (2019a). AI opportunities for healthcare must not be wasted. *Health Management*, 19.
- Floridi, L. (2019b). *The logic of information: A theory of philosophy as conceptual design* (1st edition). New York, NY: Oxford University Press.
- Floridi, L. (2019c). What the Near Future of Artificial Intelligence Could Be. *Philosophy & Technology*, 32(1), 1–15. <https://doi.org/10.1007/s13347-019-00345-y>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), 20160360. <https://doi.org/10.1098/rsta.2016.0360>
- Fong, A., Adams, K. T., Gaunt, M. J., Howe, J. L., Kellogg, K. M., & Ratwani, R. M. (2018). Identifying health information technology related safety event reports from patient safety event report databases. *Journal of Biomedical Informatics*, 86, 135–142. <https://doi.org/10.1016/j.jbi.2018.09.007>
- Fylon, F. (2015). *Joined Up Leeds*. Retrieved from Brainbox Research website: <https://www.leedscg.nhs.uk/content/uploads/2018/05/Summary-Joined-Up-Leeds-report-1.pdf>
- Garattini, C., Raffle, J., Aisyah, D. N., Sartain, F., & Kozlakidis, Z. (2019). Big Data Analytics, Infectious Diseases and Associated Ethical Impacts. *Philosophy & Technology*, 32(1), 69–85. <https://doi.org/10.1007/s13347-017-0278-y>
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumeé III, H., & Crawford, K. (2018). Datasheets for Datasets. *ArXiv:1803.09010 [Cs]*. Retrieved from <http://arxiv.org/abs/1803.09010>
- Georgiou, A., Magrabi, F., Hyponen, H., Wong, Z. S.-Y., Nykänen, P., Scott, P. J., ... Rigby, M. (2018). The Safe and Effective Use of Shared Data Underpinned by Stakeholder Engagement and Evaluation Practice. *Yearbook of Medical Informatics*, 27(1), 25–28. <https://doi.org/10.1055/s-0038-1641194>
- Gonzalez-Polledo, E. (2018). Can digital health save democracy? Meeting the cosmopolitical challenge of digital worlds. *Journal of Social and Political Psychology*, 6(2), 631–643. <https://doi.org/10.5964/jsp.p.v6i2.939>
- GP at hand. (2019). *babylon GP at hand: Our NHS Services*. Retrieved from <https://www.gpathand.nhs.uk/our-nhs-service>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, 51(5), 1–42. <https://doi.org/10.1145/3236009>
- He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1), 30–36. <https://doi.org/10.1038/s41591-018-0307-0>
- Holland, S., Hosny, A., Newman, S., Joseph, J., & Chmielinski, K. (2018). The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards. *ArXiv:1805.03677 [Cs]*. Retrieved from <http://arxiv.org/abs/1805.03677>
- Kemper, J., & Kolkman, D. (2018). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 1–16. <https://doi.org/10.1080/1369118X.2018.1477967>

- Knoppers, B. M., & Thorogood, A. M. (2017). Ethics and big data in health. *Current Opinion in Systems Biology*, 4, 53–57. <https://doi.org/10.1016/j.coisb.2017.07.001>
- Kroll, J. A., Huey, J., Barocas, S., Felten, E., Reidenberg, J., Robinson, D., & Yu, H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review*, 165.
- Krutzinna, J., Taddeo, M., & Floridi, L. (2018). Enabling Posthumous Medical Data Donation: An Appeal for the Ethical Utilisation of Personal Health Data. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-018-0067-8>
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philosophy & Technology*, 31(4), 611–627. <https://doi.org/10.1007/s13347-017-0279-x>
- Lewis, T. (2006). Seeking health information on the internet: Lifestyle choice or bad attack of cyberchondria? *Media, Culture & Society*, 28(4), 521–539. <https://doi.org/10.1177/0163443706065027>
- Macrae, C. (2019). Governing the safety of artificial intelligence in healthcare. *BMJ Quality & Safety*, bmjqs-2019-009484. <https://doi.org/10.1136/bmjqs-2019-009484>
- Mann, S. P., Savulescu, J., & Sahakian, B. J. (2016). Facilitating the ethical use of health data for the benefit of society: Electronic health records, consent and the duty of easy rescue. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083). <https://doi.org/10.1098/rsta.2016.0130>
- Michie, S., Fixsen, D., Grimshaw, J. M., & Eccles, M. P. (2009). Specifying and reporting complex behaviour change interventions: The need for a scientific method. *Implementation Science*, 4(1), 40, 1748-5908-4-40. <https://doi.org/10.1186/1748-5908-4-40>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 220–229. <https://doi.org/10.1145/3287560.3287596>
- Nag, N., Pandey, V., Oh, H., & Jain, R. (2017). Cybernetic Health. *ArXiv:1705.08514 [Cs]*. Retrieved from <http://arxiv.org/abs/1705.08514>
- Nag, N., Pandey, V., Putzel, P. J., Bhimaraju, H., Krishnan, S., & Jain, R. C. (2018). Cross-Modal Health State Estimation. *2018 ACM Multimedia Conference on Multimedia Conference - MM '18*, 1993–2002. <https://doi.org/10.1145/3240508.3241913>
- NHS Digital. (2019). National data opt-out. Retrieved from <https://digital.nhs.uk/services/national-data-opt-out-programme>
- NHS England. (2013). Care.Data. Retrieved from <https://www.england.nhs.uk/2013/10/care-data/>
- NHS England. (2018). Empower the Person: Roadmap for Digital Health and Care Services. Retrieved from <https://indd.adobe.com/view/119c9ee5-6acb-4f52-80c2-d44fc03fdc91>
- NHS England. (2019a). *NHS Diagnostic Waiting Times and Activity Data*. Retrieved from <https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2019/02/DWTA-Report-December-2018.pdf>
- NHS England. (2019b). *The NHS Long Term Plan*. Retrieved from NHS website: <https://www.longtermplan.nhs.uk/wp-content/uploads/2019/01/nhs-long-term-plan.pdf>
- NHS.uk. (2018). Can I access the medical records (health records) of someone who has died? Retrieved from <https://www.nhs.uk/common-health-questions/nhs-services-and-treatments/can-i-access-the-medical-records-health-records-of-someone-who-has-died/>
- NHSx. (2019). What we do. Retrieved from <https://www.nhsx.nhs.uk/what-we-do>
- Nuffield Council of Bioethics. (2015). *The collection, linking and use of data in biomedical research and health care: ethical issues*. Retrieved from http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf
- O'Doherty, K. C., Christofides, E., Yen, J., Bentzen, H. B., Burke, W., Hallowell, N., ... Willison, D. J. (2016). If you build it, they will come: Unintended future uses of organised health data collections Donna Dickenson, Sandra Soo-Jin Lee, and Michael Morrison. *BMC Medical Ethics*, 17(1). <https://doi.org/10.1186/s12910-016-0137-x>
- Office for National Statistics. (2019). *Avoidable mortality in the UK: 2017*. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/avoidablemortalityinenglandandwales/2017>
- Okun, S., & Wicks, P. (2018). DigitalMe: A journey towards personalized health and thriving. *BioMedical Engineering OnLine*, 17(1), 119. <https://doi.org/10.1186/s12938-018-0553-x>
- Owens, J., & Cribb, A. (2017). 'My Fitbit Thinks I Can Do Better!' Do Health Promoting Wearable Technologies Support Personal Autonomy? *Philosophy & Technology*. <https://doi.org/10.1007/s13347-017-0266-2>
- Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, 55(9), 1059–1072. <https://doi.org/10.1080/00140139.2012.691554>
- Petrini, C. (2015). On the 'pendulum' of bioethics. *Clinica Terapeutica*, 166(2), 82–84. <https://doi.org/10.7417/CT.2015.1821>
- Ploug, T., & Holm, S. (2016). Meta consent - A flexible solution to the problem of secondary use of health data. *Bioethics*, 30(9), 721–732. <https://doi.org/10.1111/bioe.12286>

- Powell, J., & Deetjen, U. (2019). Characterizing the Digital Health Citizen: Mixed-Methods Study Deriving a New Typology. *Journal of Medical Internet Research*, 21(3), e11279. <https://doi.org/10.2196/11279>
- Quinn, P. (2017). The Anonymisation of Research Data - A Pyrrhic Victory for Privacy that Should Not Be Pushed Too Hard by the eu Data Protection Framework? *European Journal of Health Law*, 24(4), 347–367. <https://doi.org/10.1163/15718093-12341416>
- Rahwan, I. (2018). Society-in-the-Loop: Programming the Algorithmic Social Contract. *Ethics and Information Technology*, 20(1), 5–14. <https://doi.org/10.1007/s10676-017-9430-8>
- Rebhan, M. (2017). Towards a systems approach for chronic diseases, based on health state modeling. *F1000Research*, 6, 309. <https://doi.org/10.12688/f1000research.11085.1>
- Reed, C. (2018). How should we regulate artificial intelligence? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170360. <https://doi.org/10.1098/rsta.2017.0360>
- Sethi, N., & Laurie, G. T. (2013). Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together. *Medical Law International*, 13(2–3), 168–204. <https://doi.org/10.1177/0968533213508974>
- Silverman, B. G., Hanrahan, N., Bharathy, G., Gordon, K., & Johnson, D. (2015). A systems approach to healthcare: Agent-based modeling, community mental health, and population well-being. *Artificial Intelligence in Medicine*, 63(2), 61–71. <https://doi.org/10.1016/j.artmed.2014.08.006>
- Sterckx, S., Rakic, V., Cockbain, J., & Borry, P. (2016). “You hoped we would sleep walk into accepting the collection of our data”: controversies surrounding the UK care.data scheme and their wider relevance for biomedical research. *Medicine, Health Care and Philosophy*, 19(2), 177–190. <https://doi.org/10.1007/s11019-015-9661-6>
- Taylor, L., Floridi, L., & van der Sloot, B. (Eds.). (2017). *Group privacy: New challenges of data technologies*. Switzerland: Springer.
- Townend, D. (2018). Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? *Human Genetics*, 137(8), 657–664. <https://doi.org/10.1007/s00439-018-1924-x>
- Turilli, M., & Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, 11(2), 105–112. <https://doi.org/10.1007/s10676-009-9187-9>
- Understanding Patient Data. (2019). How do people feel about the use of data? Retrieved from <https://understandingpatientdata.org.uk/how-do-people-feel-about-use-data>
- Vayena, E., Tobias, H., Afua, A., & Allesandro, B. (2018). Digital health: Meeting the ethical and policy challenges. *Swiss Medical Weekly*, 148(34). <https://doi.org/10.4414/sm.w.2018.14571>
- Verbeek, P.-P. (2009). Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology. *NanoEthics*, 3(3), 231–242. <https://doi.org/10.1007/s11569-009-0077-8>
- Watson, D. S., Krutzinna, J., Bruce, I. N., Griffiths, C. E., McInnes, I. B., Barnes, M. R., & Floridi, L. (2019). Clinical applications of machine learning algorithms: Beyond the black box. *BMJ*, 1886. <https://doi.org/10.1136/bmj.1886>
- Winfield, A. F. T., & Jirotko, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180085. <https://doi.org/10.1098/rsta.2018.0085>
- Woolley, J. P. (2019). Trust and Justice in Big Data Analytics: Bringing the Philosophical Literature on Trust to Bear on the Ethics of Consent. *Philosophy & Technology*, 32(1), 111–134. <https://doi.org/10.1007/s13347-017-0288-9>